

Statistical Disaggregation of Earth Science Data

Robin D. Morris, Vadim Smelyanskiy, J. Dungan, J. Coughlan NASA Ames

Goal:

- Develop statistical methodology and practice to produce dense estimates of ecological parameters that are difficult to measure directly.

Objectives:

- For ecological parameters that it is difficult to measure directly on a dense grid, develop statistical methods for estimating these parameters based on sparse measurements of the parameters, and dense measurements of covariate parameters.
- Encode the domain knowledge of the Earth Scientist into the statistical relationship between the parameters.

Technical Approach:

- Apply Bayesian methodology to the disaggregation problem.
- Estimate uncertain relationships between ecological parameters from published fieldwork data.
- Produce uncertainty estimates of the dense parameter estimates.

Technical Significance & NASA Relevance:

- NASA collects huge databases of Earth Science data. The results of this research will enable this data to be used for a wider range of Earth Science problems.
- Produce high-resolution soil water maps over large areas, to enable more accurate biospheric modeling and prediction.
- Introduce advanced statistical ideas and methodology to parts of the Earth Science community, to enable better understanding of the strengths and limitations of the data being collected and analysed.

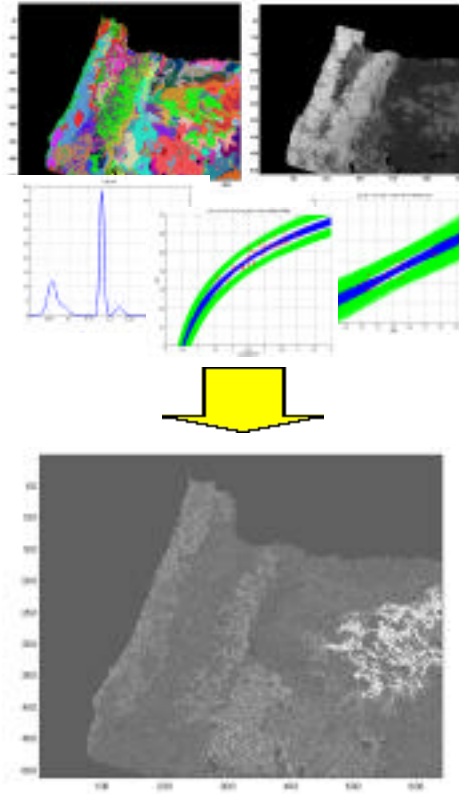
Accomplishments & Preliminary Findings:

- Completed an initial implementation of the estimation methodology.
- Applied this to a selected study area in eastern Oregon.
- Identified areas where more ecological knowledge is required to reduce the uncertainty on the estimates and to reconcile results with published information.

Personnel:

- In addition to the personnel listed above, Ms Monika Mellem contributed to the initial analysis under the NASA Undergraduate Student Research Program, Summer 2001

Technical Problem Statement



- From satellite measurements (NDVI from AVIRIS), and region-based soil databases, estimate high-resolution soil characteristics.

- Soil data (eg Available Water Capacity (AWC)) is difficult to measure over large regions. Soil maps are made by assigning point measurements to polygonal regions.

- NDVI can be measured over large areas.

- Ecological theory and measurements give statistical relationships between AWC and LAI, and LAI and NDVI.

- Bayesian statistics gives the best estimate of AWC over dense grids using the soil and LAI data, and the relationships between LAI and AWC and NDVI and LAI.

References:

- "Combining Data with Uncertain Relations", presented at the "Combating Uncertainty with Fusion" workshop, Woods Hole, MA, April 2002
- "High Resolution Soil Water from Regional Databases and Satellite Images", to be presented at the "Image Processing and Related Mathematics" workshop, Moscow State University, July 2002.

Statistical Modelling, Forecasting & Control of Large Deviation Events

Vadim Smelyanskiy, Robin Morris, NASA Ames, Dmitri Luchinskii, Lancaster (UK)

Goal: To predict catastrophic events and suggest means of causation using dynamic modeling and data mining.

Objectives:

- Modeling, forecasting and control of large occasional fluctuations leading to catastrophic behaviour in dynamical systems.
- To use Bayesian inference to learn the characteristics of the dynamical system description of ecological and biological systems.
- To use these models to predict future rare events, and to uncover correlations between the individual parameters during large fluctuational events that are strongly predictive of catastrophic behaviour.

Key Innovation:

- The concept of **optimal fluctuational paths**, along which the system moves with overwhelming probability during a large fluctuational event.
- A new algorithm for learning the dynamical system model by analysis of the return paths from rare, large deviations, rather than the conventional approach of disregarding large fluctuations as “outliers”.
- A new algorithm for model identification in periodically driven systems.
- Characterisation of the learning rate for this algorithm, and the uncertainties in the estimates of the dynamical system model parameters.
- The ability to predict, from the learned model, when the system begins to depart on a large fluctuation event.

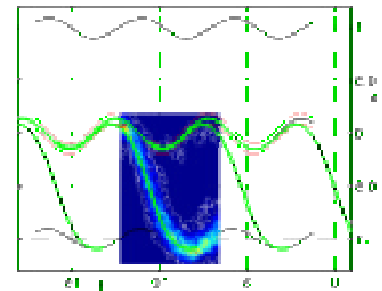
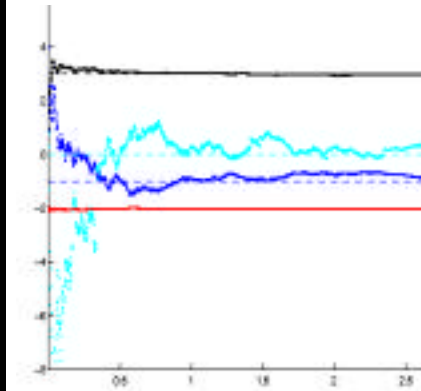
Accomplishments & Preliminary Findings:

- A detailed study was performed of climatic, tree type, fire ignition event frequency and extent of subsequent fire to learn a model of forest fire behaviour. Due to limited data availability, it was not possible to collect sufficient statistics for similar areas to reliably infer the model
- Experiments with an analogue nonlinear dynamical system have shown that the proposed algorithm can correctly infer the system characteristics, and that the theoretical learning rates are observed in practice.
- Data has been collected for the respiratory/cardiovascular system, and a dynamical system model has been suggested.

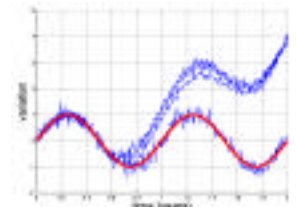
Personnel:

- Dr R. Neemani and Dr. P. Votava provided the forest fire data.

Technical Problem Statement:



- Mine from the data base those trajectories that reach the catastrophic state.
- From the return paths from the catastrophic state, infer the dynamical model.
- By monitoring the current evolution of the system we can determine whether we are travelling along a high probability tube to a catastrophic event.
- Build the multidimensional prehistory distribution for a set of parameters for the evolution paths which end at the catastrophic event.
- By analysing the distribution we hope to discover previously unknown statistical dependencies that lead up to catastrophic events.



Technical Significance & NASA Relevance:

- NASA operates and observes a wide range of systems for which large, rare fluctuations have important consequences, eg **catastrophic events in ecological systems** (eg large forest fires, floods); **human factors** (eg heart failure due to subtle changes in cardiovascular/respiratory dynamics caused by spaceflight); **climate forecasting** (eg switching behaviour in the Bearing Sea);

References:

- “Inference of Noisy Dynamical Systems Using Large Deviations in the Data”, Technical Report, NASA Ames, 2002.